

Assessing Growth in Teaching Knowledge

Susan E. Schultz¹
Stanford University

The ability to transform subject matter knowledge requires more than knowledge of the substance and syntax of one's discipline; it requires knowledge of learners and learning, of curriculum and context, of aims and objectives, of pedagogy. (Grossman, Wilson, & Shulman, 1989, p 32).

As researchers recognize the complex, multi-faceted nature of teacher knowledge, their focus has changed from reporting on what teachers do to examining what they know and how they use that knowledge to inform their teaching. In the late 1980s, research began to focus on identifying various aspects of a knowledge base necessary for successful teaching (Ayers, 1988; Reynolds, 1989; Shulman, 1987). Building on this work, Darling-Hammond, Wise, and Klein (1999) posited a new approach to teacher assessment for licensing based on a schema for a teaching knowledge base. Their schema focuses on the use of specialized teaching knowledge that is built upon subject matter and liberal arts knowledge and includes knowledge about: (1) learners and learning; (2) curriculum and teaching; and (3) contexts and social foundations of education. This knowledge, they argued, is applied in situations where students are diverse, goals are multidimensional, and judgment is always needed.

Teacher education programs embracing this kind of schema face the challenge of revising their curriculum and developing assessment tools for evaluating the knowledge base of pre-service candidates. In the past,

Susan E. Schultz is a Social Science Research Associate and Lecturer in the School of Education at Stanford University.

evaluation of teacher performance often focused on generic behaviors related to common instructional strategies (Medley, Coker, & Soar, 1984). More recently, educational researchers and test developers have focused more on teachers' subject matter knowledge as tested through tools like the Praxis and California's Single Subject Assessment Tests. Most states now test basic skills and subject matter knowledge (Madaus & Mehrens, 1990). However, useful assessments of teachers' knowledge of development, learning, and pedagogical strategies have been more difficult to develop. To address this gap, the Interstate New Teacher Assessment and Support Consortium (INTASC, 1992), in conjunction with the Educational Testing Service [ETS], developed a prototype for a Test for Teaching Knowledge [TTK]. The TTK is designed to evaluate the foundational teaching knowledge of learning, learners, and development described by Darling-Hammond, Wise, and Klein (1999) and is intended to be used as a measure of candidates' teaching readiness.

This study focuses on the Stanford Teacher Education Program's (STEP) use of the INTASC Test of Teaching Knowledge to gauge changes in preservice candidates' areas of knowledge and skills over the twelve-month program. This paper describes the test, reports changes in the pre-service candidates' performance on the assessment, and discusses implications of the findings for the STEP program and other teacher education programs.

Using the pilot version of the test to assess preservice candidates' teaching knowledge also allowed us to investigate important questions about the validity of the test and scoring rubric. Based on our findings, we believe that these are important questions that need further study. Since some of the development work on the TTK may inform the development of the Teaching Performance Assessment ETS is currently working on for California's Commission on Teacher Credentialing, these issues may have current and future significance for performance assessment of teaching in California and elsewhere.

The Interstate New Teacher Assessment and Support Consortium

The Interstate New Teacher Assessment and Support Consortium (INTASC, 1992) is a group of more than 30 state education agencies and professional education organizations that share the goal of reforming the education, licensing, and on-going professional development of teachers. The basic premise of INTASC is that an effective teacher must be able to integrate content knowledge with pedagogical understanding to assure that all students learn and perform at high levels. INTASC promotes this vision through the development of model standards and assessments for

beginning teachers. The INTASC Standards are organized around principles of teaching and serve as a framework for the systemic reform of teacher preparation and professional development. The ten core INTASC principles, which serve as the basis for standards for beginning teachers since developed in separate subject areas, are as follows:

Principle #1: The teacher understands the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and can create learning experiences that make these aspects of subject matter meaningful for students.

Principle #2: The teacher understands how children learn and develop, and can provide learning opportunities that support their intellectual, social and personal development.

Principle #3: The teacher understands how students differ in their approaches to learning and creates instructional opportunities that are adapted to diverse learners.

Principle #4: The teacher understands and uses a variety of instructional strategies to encourage students' development of critical thinking, problem solving, and performance skills.

Principle #5: The teacher uses an understanding of individual and group motivation and behavior to create a learning environment that encourages positive social interaction, active engagement in learning, and self-motivation.

Principle #6: The teacher uses knowledge of effective verbal, nonverbal, and media communication techniques to foster active inquiry, collaboration, and supportive interaction in the classroom.

Principle #7: The teacher plans instruction based upon knowledge of subject matter, students, the community, and curriculum goals.

Principle #8: The teacher understands and uses formal and informal assessment strategies to evaluate and ensure the continuous intellectual, social and physical development of the learner.

Principle #9: The teacher is a reflective practitioner who continually evaluates the effects of his/her choices and actions on others (students, parents, and other professionals in the learning community) and who actively seeks out opportunities to grow professionally.

Principle #10: The teacher fosters relationships with school colleagues, parents, and agencies in the larger community to support students' learning and well-being. (INTASC, 1992)

More than 30 states have adopted or adapted these principles to guide their licensing and program approval decisions in teacher education. California's Standards for the Teaching Profession [CSTP] (Califor-

nia Department of Education, 1997) are closely compatible with the INTASC principles. To evaluate these standards, the consortium is developing assessments focused on both teaching knowledge and teacher performance that are designed to supplement existing subject matter tests. The first assessment, called the INTASC Test for Teaching Knowledge, seeks to evaluate preservice candidates' knowledge of learners, learning, and core elements of teaching. STEP administered the pilot Test of Teaching Knowledge as a pre- and post-test to candidates in the classes of 2000 and 2001.

The second phase of evaluation is a performance-based portfolio assessment that aims to capture the beginning teacher's teaching skills in the context of a classroom where they are embedded in subject matter. INTASC has developed a prototype for this portfolio that has been most fully implemented in Connecticut. The portfolio resembles the STEP portfolio in many respects.

Test of Teaching Knowledge

The Test of Teaching Knowledge may be used by states as a test for initial licensure or by colleges as a formative or summative assessment near the end of the preservice experience. The pilot test contained twenty-six constructed response items to evaluate a preservice candidate's professional knowledge in areas such as child development, theories of teaching and learning, diagnostic and assessment skills, the role of student background in the learning process, and other foundational knowledge and skills. The test developers expect candidates to complete the assessment within four hours. The twenty-six constructed response items on the pilot test used in this study are distributed across four sections. In the first section, candidates respond to four multiple part questions addressing specific knowledge about learners and how that knowledge might influence the learning and/or teaching process. The second section asks candidates to read a case study or classroom vignette focusing on aspects of learning, student behavior, or classroom instruction and to answer seven questions related to the case study. The third section provides a "folio" or a collection of documents and asks candidates to answer seven questions dealing with a particular learner or aspect of learning or teaching illustrated in the documents. In the final section, candidates answer eight short, focused questions assessing propositional knowledge about specific theories, learning needs, instructional strategies, or teaching concepts.

The distribution of test items by knowledge constructs for the pilot version of the test is presented in Table 1. Nine test items centered on

**Table 1:
Distribution of Test Items by Knowledge Constructs**

<u>Types of Knowledge</u>	<u>Test Items</u>
Knowledge about Learners / Learning	9*
Knowledge about Curriculum & Teaching	20*
Knowledge about Contexts & Foundations of Education	7**

*3 test items assessed knowledge about learners as well as curriculum and teaching

** 7 test items evaluating knowledge about context and foundations overlapped with items in the other two categories.

knowledge about learners and learning, assessing knowledge of human growth and development and learning theory. Twenty items assessed preservice candidates' knowledge about curriculum and teaching, focusing on instructional design, pedagogical strategies, assessment techniques, language acquisition, and communication skills. Three test questions overlap these categories, including both specific knowledge of the learner and of how that knowledge impacts instruction. Seven items assessed knowledge about contexts and foundations of education as well as knowledge about learners or curriculum and teaching. The test items do not measure preservice candidates' subject matter knowledge. In a few of the test questions, the prompt requires candidates to respond using examples from their subject specific content domain, but the focus is evaluating candidates' curriculum and teaching knowledge.

The specific items on the pilot form of the test have not been released for public use. Sample items are available on the INTASC website; one example item is provided here for the purpose of explaining the task demand, the candidate response, and the scoring system.

Task: This sample item seeks to measure a preservice candidate's ability to understand how students differ in their approaches to learning and creates instructional opportunities that are adapted to diverse learners (INTASC, Principle # 3).

Prompt: What characteristics or behaviors would help you identify a student who has a strong bodily-kinesthetic intelligence? Provide one example of an instructional strategy that would take advantage of such a student's strength. Base your response on principles of multiple intelligences and principles of providing instructional opportunities for diverse learners.

Response: Preservice candidates read the prompt, suggest a specific instructional strategy to help identify a student with strong bodily-kinesthetic intelligence, and explain the connection between the instructional strategy and the principles of multiple intelligences.

Scoring: For twenty-two of the test items, a 3-point scoring rubric (0 indicating no response or an incorrect response to 2 representing an excellent response) was developed by ETS and used for scoring. On the remaining four test items, a 2-point scoring rubric (0 indicating no response or an incorrect response to 1 representing a correct response) was used. The written response of the preservice candidate was compared with the possible answers on the rubric and a numerical value was assigned to the response.

Methodology

In this study a pilot version of the test was administered as a pre and post measure of teaching knowledge at the beginning and the end of the twelve-month teacher education program in 1999-2000. All STEP candidates (N=59) completed the pre-test but, since it was not a requirement in that year, only 25 candidates completed the post administration of the test. These 25 candidates earned undergraduate diplomas from highly competitive institutions with an average G.P.A. of 3.46. Twenty percent had taught prior to entering STEP.

For both the pre- and post-test administrations each candidate received a test booklet and the option of a pencil or pen to write his/her answers. All candidates finished the test within the four-hour time frame. The pilot rubrics developed by ETS were used to score the items. The low-inference approach used in the rubrics supported high reliability in scoring, although as discussed later, it may have compromised the capacity of the scoring system to detect important qualitative differences in responses. The small sample size and potential for selectivity bias suggest cautious interpretation of study findings.

Results and Discussion

To detect changes in candidates' teaching knowledge, the initial analysis of data examined the change in scores between the pre- and post-test for each candidate (see Table 2).

On thirteen items, 96% of the candidates earned a 2-point change and 4% showed a 1-point change in score between the pre- and post-test. All candidates reached the maximum score on the post-test. On six items, all candidates (100%) showed a gain score of 1-point, with 15% reaching the maximum score on the post-test. On seven items, most candidates (92%) revealed no change in score, whereas 8% earned a 1-point change. All candidates reached the maximum score on the post-test. The lack of change was due to high scores on the pre-test, a signal to us to examine

Table 2
Percentage of Candidates by Change in Score by Item Frequency

Number of Items (N = 26)	No Change in Score	Partial Change		Maximum Change*
	Pre=Post Score	Pre=0 to Post=1 pt	Pre=1 to Post=2 pts	Pre=0 to Post=2
13	0 %	0 %	4 %	96 %
6	0 %	85 %	15 %	0 %
7	92 %	0 %	8 %	0 %

*On 4 items, the rating scale was only a 0-1 scale. These were also coded as "full credit" where the pre-test score was 0 and the post-test score was 1.

the nature of the items themselves. If candidates could answer these items correctly before they entered preservice preparation, we reasoned, these might not be particularly powerful assessments of professional teaching knowledge. We clustered these items by the degree of change in scores for purposes of further analysis.

Identifying the nature of the knowledge contained within the items clustered together in the previous analysis was the next step in the analysis. Table 3 shows the distribution of content knowledge being assessed in the item clusters.

Items with Large Gain Scores

Candidates showed the largest gain score between the pre- and post-test on the first cluster of thirteen items. This cluster focuses primarily on knowledge about curriculum and teaching with two items about the knowledge of learners and learning. In these items, candidates need to use their knowledge to examine the appropriateness of the lesson for the stated goals, discuss the effectiveness of specific pedagogical strategies, critique assessment strategies in light of the lesson objectives, evaluate interactions between the teacher and students, and apply principles of motivation to enhance student learning.

Table 3
Distribution of Content Domains by Item Cluster

Items (N = 26)	Knowledge about Learners and Learning		Knowledge about Curriculum and Teaching				
	Emotional, Social, & Cognitive Development Differences	Learning Theory & Learning Differences	Motivation & Behavior	Instructional Design	Pedagogical Strategies	Assessment Techniques	Language & Communication
13	0	0	2	3	4	2	2
6	2	2	0	0	1	1	0
7	0	0	0	3	3	1	0

Two responses illustrate the growth in candidates' knowledge between the pre- and post-test. In the first example, the test item asks candidates to answer a question basing their response on the principles of motivation and the learning process (INTASC, Principle 5).

Pre-Test: "I have not studied the principles of motivation and the learning process." (ID # 5, 6/99)

Post-Test: "While Mrs. Chen is certainly helping build self esteem in her students through positive comments, she is not giving students much feedback on how they might improve. Thus, there is no way for students to monitor the quality of their own work. If Ms. Chen were to give positive reinforcement along with specific constructive criticism, then her students would know how to measure their own progress and be motivated to improve. This would also foster the incremental model of intelligence which is based on the idea that achievement is the result of effort rather than intelligence." (ID # 5, 6/00)

In the pre-test response the candidate does not offer any response other than to explain that she has not studied these concepts. Other candidates wrote similar responses or left the paper blank on these items during the pre-test. The post-test response demonstrates the candidate's knowledge of motivation theory and principles of feedback. She points out that the teacher is building students' self esteem through the use of positive comments but identifies the importance of providing specific feedback to promote improvement in student performance. She discusses how letting students know what needs improving will result in higher motivation levels and eventually enable them to monitor their own work. She also identifies the influences of feedback and opportunities for self-assessment on students' views of their own intelligence, which may influence their academic self-concept and future willingness to put forth effort.

Another example demonstrates the change in the quality of the candidate's knowledge between the pre-test and post-test for a score change of 1-point. For this test item, candidates read a prompt and respond to a question about small group instruction (INTASC, Principle 4).

Pre-Test: "Small group work enables many students to participate at a time in active learning and discussion." (ID # 22, 6/99)

Post-Test: "You could explain the research that shows the relationship between student participation in group work activities and student learning gains. Students who participate in asking and answering questions or in class activities and discussion show greater learning gains than those who do not. Group work enables more students at a time to engage in the class activities and discussions or in asking/ answering questions." (ID # 22, 6/00)

In the pre-test response the candidate makes a very general statement about the advantage of using group work as an instructional strategy by providing opportunities for students to participate in the learning process. In contrast, in the post-test response the candidate describes research connecting student participation in group work tasks and student learning gains. Although the candidate does not cite specific research the students studied (e.g., Cohen & Lotan, 1997; Schultz, 1999), s/he explains key findings of the research and describes the student activities during group work instruction that promote higher learning gains.

When examining the candidates' gain scores and the types of knowledge on these items, some clear patterns emerge regarding the kind of content candidates have mastered. Candidates are able to critique a lesson plan and suggest concrete changes in content, sequence, and assessment to match the stated learning goals. Given a specific pedagogical strategy, candidates can reflect on the appropriateness of the strategy to the stated objectives and discuss the potential impact on student learning.

To evaluate the relationship of the test to the STEP curriculum and to gauge the effectiveness of courses whose content overlapped the test, the distribution of content on the test items was compared with the content of STEP courses. The knowledge base representing these thirteen items is the primary focus of the subject-specific Curriculum and Instruction course. In addition, the Principles of Learning for Teaching course, the Adolescent Development course, and a course on Teaching and Learning in Heterogeneous Classrooms cover important areas of tested content. In the three-quarter Curriculum and Instruction sequence, candidates explore the key concepts on the content domain; plan a two-week unit including the development of instructional tasks and selection of appropriate teaching techniques; develop assessment strategies to evaluate students' learning and guide instruction; teach the unit in their school placements; and revise the unit. In the Principles of Learning course, learning theories are examined and concepts like scaffolding, assisted performance, strategies to assist cognitive processing, feedback and assessment, and strategies for culturally responsive teaching are explored. Adolescent Development examines social, emotional, and cognitive development, motivation and behavior, and features of home, school, and community contexts that influence development. Heterogeneous Classrooms addresses cooperative learning strategies, teaching of English language learners, and supports for teaching to multiple learning styles. Among them, these courses cover most of what is included in the portions of the TTK that showed the most growth for STEP teachers.

Items Showing Some Growth

Among the second cluster of items, four emphasize knowledge about learners and learning and two items center on knowledge about curriculum and teaching. In these six items, candidates share their knowledge about cognitive development, exceptional learner needs, how specific information about the learner might influence student learning and/or guide teaching practices; and strategies for developing and evaluating assessment techniques.

All of the candidates showed partial gains between the pre- and post-test on these six items. About 85% progressed from a “0” to a “1” and about 15% progressed from a “1” to a “2” in their scores on the items. In these items, candidates answer multiple parts and need to generalize beyond the content to broader implications for teaching and/or student learning. In this item, candidates review a situation and discuss assessment strategies and implications for teaching (INTASC, Principle 8).

Pre-Test: No response (ID # 6, 6/99)

Post-Test: “(The teacher) could assess the group project by using a rubric on which she assesses the quality of content learning, group effort, oral presentations, and visuals (like a chart). Rubrics are useful because you can break down the assessment into different areas, highlighting a student’s strength and areas for improvement.” (ID # 6, 6/00)

In the pre-test, the candidate left the page blank. Leaving the response completely blank or writing “I have no idea” or “I’m looking forward to learning about this during my year at STEP” were common responses on the pre-test for these 6 items. The candidate’s response in the post-test discusses the use of a rubric as an assessment tool and how the rubric could be used to measure a variety of indicators of the group’s performance. The benefits of using a rubric as a diagnostic tool for both the student and the teacher are also highlighted.

Although there was considerable improvement in the response on the post-test as compared to the blank response on the pre-test, the candidate fails to address how this assessment strategy would shape or affect his or her teaching. Thus, this response earned a partial credit (1 point) instead of full credit. All six of these items required candidates to reflect on how the specific topic would influence an aspect of teaching and/or how the information would impact student learning.

In each case, candidates’ responses clearly indicate an understanding of the content requested in the question, but they less frequently discuss how they would apply that specific content to instructional practices. For example, candidates were able to identify characteristics

of learners, ranging from stages and indicators of social and cognitive development to identification of specific learning disabilities, but they did not fully explain how knowing this information would influence their teaching and/or student learning in specific ways.

Knowledge about learners, learning, and assessment resides in a series of foundation courses within the STEP curriculum, especially Adolescent Development and Principles of Learning for Teaching. These courses expose candidates to a range of content centered on knowing more about learners and the learning process and about how to design instruction. The fact that candidates are able to articulate concrete information and show an awareness of the multiple dimensions of learners is encouraging. However, more sophisticated questions about how to design and validate assessments or shape instruction to meet the needs of students with specific learning disabilities was more difficult for most candidates, in part because of limited instructional depth in these areas and in part because of the lack of opportunity to apply these state-of-the-art ideas in many current student teaching placements.

STEP is addressing this challenge in three ways: (1) deepening course content around assessment design and around the specific teaching strategies appropriate to students with special needs; (2) requiring candidates to think about how different characteristics of learners might influence the instructional design of the class, and (3) inviting cooperating teachers to attend STEP classes where these issues are explored and developing inservice learning opportunities so that applications in the classroom become more widespread.

Items That Did Not Measure Growth

The third cluster, containing seven items, concentrates on knowledge about curriculum and teaching. In most of these items, candidates answer questions pertaining to a case study that describes a lesson and includes interactions between the teacher and students in a classroom setting. In answering these questions, candidates were asked to critique the effectiveness of a lesson, identify classroom management strategies, analyze the appropriateness of instructional strategies, and evaluate interactions between the teacher and students.

Surprisingly, analysis of these seven items reveals that most candidates earned full credit on both the pre- and post-test. The focus of our analysis switched from candidates' growth in teaching knowledge to posing questions about these test items. As a preface to the findings, it is essential to acknowledge that we were using the pilot versions of the test items and scoring rubrics for the INTASC Test for Teaching Knowl-

edge. The test items as well as the scoring rubrics may have been significantly altered since the pilot version.

A closer examination of these items indicates two types of problems: (1) test items that did not capture professional knowledge of teaching; and (2) scoring rubrics that did not capture differences in the quality of candidates' responses. An example and discussion of both categories of items follows.

Failure To Tap Professional Teaching Knowledge. Three of the seven items categorized as not tapping professional knowledge of teaching activate a response derived from personal experience and common sense rather than an assessment of specific knowledge. The errors teachers make in the scenarios are painted as so obvious and egregious that they require no sophisticated analysis. These items elicit a similar candidate response regardless of whether they are given at the beginning or end of the teacher education program. The case study and prompts in these items are situations that most students encounter at one time or another while in school. For this item, candidates read an interaction between the teacher and two students in terms of motivation and adolescent development.

Pre-Test: "The significance of Mrs. H.'s response to Bette and KT is profound. First, Mrs. H., by being inconsistent, creates unclear standards and expectations. Such an approach decreases adolescent motivation. In addition, it creates a situation in which it is difficult for students to succeed. Bette, by not being held accountable for her late work, learns that she can give an excuse and not have consequences. KT, in watching the interaction learns that Mrs. H. will be biased and inconsistent. In both cases, the students are not observing a positive role model. Lack of positive adult models contributes to a lack of youth motivation and development." (ID # 4 , 6/99)

Post-Test: "The text portrays extreme significance regarding Mrs. H.'s inconsistency, morality, and motivation strategies. KT learned in this incident that Mrs. H. does not treat people fairly. In that morality is a large part of adolescent development, it is unfortunate that one of KT's teachers is demonstrating unethical and unfair practices. KT, in trying to figure out what is right and wrong in the world learns that whining, lying, etc leads to a better grade. This message is confusing when aligned to fairness & equity in other classes with life examples. In addition, KT learns that, no matter what she does, Mrs. H. will be subjective in her evaluation. This, in turn, leads to KT feeling unmotivated because there is no clear, consistent standard for her to strive toward." (ID # 4, 6/00)

This example demonstrates how the prompt activates a response

addressing issues of fairness and teacher inconsistency rather than specialized teaching knowledge. In the post-test, the candidate uses expanded vocabulary and hints of a larger knowledge base about motivation and adolescent development, but the core of her comments parallels her response in the pre-test. Other candidates' responses to this item included accounts of personal experiences with former teachers remembered as inconsistent and unfair in the assignment of either punishment or grades.

No Differences Detected by Rubric. In the other four items, the differences in the quality of the candidate's response between the pre- and post-test were not captured by the scoring rubric, which gave both responses top scores. To understand this category of items, review the following example. The test item directs the candidate to first read a case study and to then reflect on the lesson in light of the principles of effective instructional planning (INTASC, Principle 7).

Pre-Test: "I think the video could have been used more effectively if some discussion about the purpose or goals of the video had preceded it. It would help the students understand what they were supposed to learn while watching the video. The teacher could also ask students what they know about the topic before watching the video." (ID # 13, 6/99)

Post Test: "If the teacher gave a 'case' that was relevant to the students & related to the videotape first, then she would have a better 'hook' for her students. This would have motivated them to pay attention to the video & perhaps begin thinking of how it applies to their own lives. Another thing she could have done is use the video as a conclusion to a discussion so that students could share their own experiences or explore what they thought about the issues on their own." (ID # 13, 6/00)

In both pre- and post-test responses, this candidate proposes that setting a context and purpose for viewing a video would facilitate students' understanding. In the pre-test response, the candidate also suggests connecting the video to the students' prior knowledge, but does not comment on the connection of the video to the rest of the lesson or to the students. In the post-test response, the candidate offers specific strategies for hooking the students and suggests the use of a case that is relevant to students and to the video. The post-test response reflects the candidate's awareness of the importance of making relevant connections between students and video content and suggests a mechanism for motivating students. The candidate also offers an alternative sequencing of the lesson. Qualitative differences in teaching knowledge demonstrated in the pre- and post-test responses were not detected by the scoring rubric, which classified both responses in the same way.

Conclusion

Candidates' teaching knowledge as measured by the INTASC Test for Teaching Knowledge improved over the course of the twelve-month teacher education program. The candidates' growth included knowledge about curriculum and teaching as well as knowledge about learners and learning. To demonstrate knowledge about curriculum and teaching, candidates identified effective learning environments, critiqued instructional materials; discussed a variety of pedagogical strategies, and reflected on ways to communicate and interact with students. In assessing knowledge about learners and learning, candidates showed evidence of being able to identify learning theories, stages of adolescent development, second language acquisition techniques, and assessment strategies. Between the pre- and post-test, STEP candidates' responses show growth in their knowledge of specific content. In some areas, they did not demonstrate a fully-developed ability to discuss the implications of that knowledge for their teaching decisions. Reflecting on these areas has led to changes in the STEP curriculum.

The Test for Teaching Knowledge provides a useful lens for evaluating program effectiveness in developing specific dimensions of teaching knowledge. When administered as a pre- and post-test, it provides a mechanism for capturing candidates' growth across specific standards of foundational knowledge for teaching. This is an important accomplishment in a field where such pedagogical knowledge tests have been routinely criticized in the past as over-simplifying teaching and failing to tap important teaching knowledge (Haertel, 1991; Darling-Hammond, Wise, & Klein, 1999). As with any pilot assessment instrument, revisions to test items and scoring rubrics would strengthen the validity of the test. Items that are easily answerable before candidates enter a teacher education program are likely measuring something other than professional knowledge. Rubrics that do not allow measurement of more sophisticated understandings of teaching and learning cannot fully capture the growth that candidates may make as they become increasingly knowledgeable. STEP is currently revising prompts and scoring rubrics to assess changes in candidates' teaching knowledge across specific dimensions of the program, just as it is revising its curriculum to better support candidate learning.

Note

¹The author wishes to thank Linda Darling-Hammond for her contributions to this paper.

References

- Ayers, W. (1988). Facts or fancy: The knowledge base quest in teacher education. *Journal of Teacher Education* 39, 24-29.
- California Department of Education. (1997). California standards for the teaching profession [CSTP]. www.cde.ca.gov. Sacramento, CA: Author.
- Cohen, E. G. & Lotan, R. A. (1997). *Working for equity in heterogeneous classrooms: Sociological theory in practice*. New York: Teacher's College Press.
- Darling-Hammond, L., Wise, A. E., & Klein, S. P. (1999). *A license to teach*. San Francisco: Jossey-Bass Inc.
- Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher*. New York: Pergamon.
- Haertel, E. H. (1991). New forms of teacher assessment. In G. Grant (Ed.), *Review of research in education*, 17, (pp. 3-29). Washington, DC: American Educational Research Association.
- Interstate New Teacher Assessment and Support Consortium [INTASC]. (1992). *Model standards for beginning teacher licensing and development: A resource for state dialogue*. Washington, DC: Council of Chief State School Officers.
- Madaus, G. & Mehrens, W. A. (1990). Conventional tests for licensure. In *The new handbook of teacher evaluation* (pp. 257-277). Newbury Park, CA: Sage.
- Medley, D. M., Coker, H., & Soar, R. S. (1984). *Measurement-based evaluation of teacher performance*. New York: Longman.
- Reynolds, M. C. (1989). *Knowledge base for the beginning teacher*. New York: Pergamon.
- Schultz, S. E. (1999). *To group or not to group: Effects of group interaction on students' declarative and procedural knowledge in science*. Unpublished doctoral dissertation, Stanford University. Stanford, CA.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review*, 57, 1-22.